

# Envelope Theorems for Non-Smooth and Non-Concave Optimization<sup>\*</sup>

Andrew Clausen<sup>†</sup>  
University of Pennsylvania

Carlo Strub<sup>‡</sup>  
University of St. Gallen

First version: June 9, 2010

This version: April 23, 2012

## Abstract

We study general dynamic programming problems with continuous and discrete choices and general constraints. The value functions may have kinks arising (1) at indifference points between discrete choices and (2) at constraint boundaries. Nevertheless, we establish a general envelope theorem: first-order conditions are necessary at interior optimal choices. We only assume differentiability of the utility function with respect to the continuous choices. The continuous choice may be from any Banach space and the discrete choice from any non-empty set.

*Keywords:* Envelope theorem, differentiability, dynamic programming, discrete choice, non-smooth analysis

---

<sup>\*</sup>We are very grateful to Guy Arie, Martin Barbie, Christian Bayer, Aleksander Berentsen, Carlos Carillo-Tudela, Harold Cole, Russell Cooper, Thomas Davoine, Jesús Fernández-Villaverde, Marcus Hagedorn, Christian Hellwig, Leo Kaas, Andreas Kleiner, Dirk Krueger, Tzuo Hann Law, Yvan Lengwiler, Alexander Ludwig, George Mailath, Steven Matthews, Guido Menzio, Nirav Mehta, Leonard Mirman, Kurt Mitman, Georg Nöldeke, Nicola Pavoni, Andrew Postlewaite, Kevin Reffett, José Rodrigues-Neto, Felipe Saffie, Florian Scheuer, Karl Schlag, Ilya Segal, Shouyong Shi, Lones Smith, Kyungchul Song, Xi Weng and Mark Wright for fruitful discussions.

<sup>†</sup>[clausen@sas.upenn.edu](mailto:clausen@sas.upenn.edu)

<sup>‡</sup>[cs@carlostrub.ch](mailto:cs@carlostrub.ch)

# 1 Introduction

Optimization problems that involve both discrete and continuous choices are common in economics. Examples include the trade-off between consumption and savings alongside the discrete decision of whether to work, accept a job offer, declare bankruptcy, go to college, or enroll children in child care.<sup>1</sup> In addition, we show how non-smooth optimization problems, such as capital adjustment in the presence of fixed costs, may be recast as mixed continuous and discrete choice problems. In the absence of lotteries or other smoothing mechanisms, such problems create kinks in the value function where agents are indifferent between two discrete choices. A second type of kink arises when constraints become binding. As a result, the value function is non-differentiable, non-concave, and may even lack directional derivatives. Can first-order conditions be applied under such circumstances?

This paper provides two general *envelope theorems*. The first relates to static optimization problems. [Figure 1a](#) illustrates an example where an investor maximizes his profit by choosing the size of his investment,  $c$ , and the product,  $d_1$  or  $d_2$ , to invest in. The investor takes the upper envelope over the two per-product profits  $f$  and maximizes it with respect to the continuous choice  $c$ . We assume  $f(\cdot, d)$  is differentiable for each discrete choice  $d \in \{d_1, d_2\}$ . Observe in the figure that the upper envelope has only downward kinks but no upward kinks. Moreover, maxima may not occur at downward kinks. Therefore, our static envelope theorem concludes that interior maxima only occur at differentiable points. In other words, at an investment level where the investor is indifferent between the two products, he strictly prefers to increase the investment and choose product  $d_1$ , or decrease the investment and choose product  $d_2$ . [Amir, Mirman and Perkins \(1991, Lemma 3.4\)](#) and [Milgrom and Segal \(2002, Corollary 2\)](#) provide special cases of this theorem under the assumptions of supermodularity and equidifferentiability, respectively.<sup>2</sup>

Our second envelope theorem applies this intuition to dynamic settings. When an agent makes both discrete and continuous choices subject to some constraint, the value function has (potentially infinitely many) kinks. In [Figure 1b](#),  $c$  represents effort, and the two curves represent the payoffs from attending college or not. As before, discrete choices may lead to downward kinks. In addition, binding constraints may lead to upward (or downward) kinks. Nevertheless, we show that at interior optimal choices, the value function is differentiable; the agent never chooses a savings level where he is indifferent between college or not.<sup>3</sup> More specifically, our theorem applies if the choice is an *optimal one-period interior choice*,

---

<sup>1</sup> [Eckstein and Wolpin \(1989\)](#), [Rust \(2008\)](#), or [Aguirregabiria and Mira \(2010\)](#) list many more examples.

<sup>2</sup> A totally different approach by [Renou and Schlag \(2009\)](#) does not study derivatives at all but uses the weaker notion of “ordients” (ordered gradients).

<sup>3</sup> In complementary work, [Rincón-Zapatero and Santos \(2009\)](#) study the differentiability of value functions at boundary choices.

which means the agent is able to increase or decrease his continuous choice today without changing any other choices. For example, this condition is met if the agent can increase or decrease his savings today without changing his college or future savings decisions.

Previous envelope theorems in dynamic settings do not accommodate discrete choices and impose additional assumptions. [Mirman and Zilcha \(1975, Lemma 1\)](#) and [Benveniste and Scheinkman \(1979, Theorem 1\)](#) impose concavity assumptions, and [Amir et al. \(1991, Lemma 3.4\)](#) assume supermodularity.

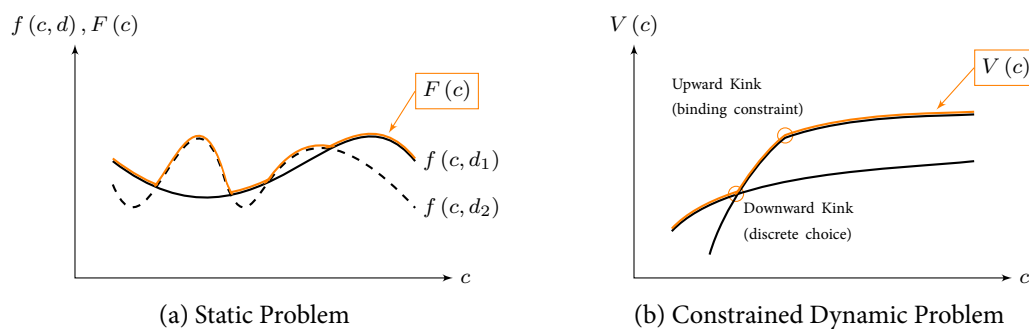


Figure 1: Mixed continuous and discrete choice problems

The concept of directional derivatives is central to the proofs of previous envelope theorems. However, in our fully general setting, directional derivatives may not exist everywhere, so a new approach is required. We apply Fréchet sub- and superdifferentials, and their one-dimensional analogues which we call Dini sub- and superdifferentials.<sup>4</sup> They capture what we think of as upward and downward kinks.

This paper is organized as follows: [Section 2](#) states our envelope theorems. All proofs go into [Section 3](#) which contains additional general lemmata on kinks and upper envelopes. There, we also discuss the relationship of our results to previous publications. [Section 4](#) illustrates the breadth of applications of our envelope theorems to non-smooth and non-concave dynamic programming problems. The proofs of the Banach space versions of our theorems are in the appendix. Nevertheless, we recommend reading them as they are more elegant (but less intuitive) than the standard versions.

<sup>4</sup> The terminology “Dini sub- and superdifferential” does not appear to be widespread. On the other hand, “Fréchet sub- and superdifferential” is a standard generalization of the convex analysis notion of “subdifferential” to non-convex functions.

## 2 Theorems

An agent makes a *continuous choice*  $c \in C$  and a *discrete choice*  $d \in D$ . Initially, we require the continuous choice set  $C$  to be a subset of  $\mathbb{R}$ ; [Appendix A](#) generalizes all theorems to allow  $C$  to be a subset of any Banach space. We allow the discrete choice set  $D$  to be any non-empty set, e.g. a finite set such as {full time work, part time work, not work}, a continuous space such as  $\mathbb{R}^2$ , or an infinite dimensional space such as  $C[0, 1]$ .

**Definition 1.** We say  $F$  is the *upper envelope* of  $\{f(\cdot, d)\}_{d \in D}$  if  $F(c) = \sup_{d \in D} f(c, d)$ .

Our static envelope theorem asserts that non-differentiable points are never optimal choices. An agent is never indifferent between two discrete choices after making an optimal continuous choice (unless the discrete choices are locally equivalent).

**Theorem 1.** Suppose  $F$  is the upper envelope of a (possibly infinite) set of differentiable functions  $\{f(\cdot, d)\}$ . If  $(\hat{c}, \hat{d})$  maximizes  $f$ , and  $\hat{c} \in \text{int}(C)$ , then  $F$  is differentiable at  $\hat{c}$  and satisfies the first-order condition  $F'(\hat{c}) = f_c(\hat{c}, \hat{d}) = 0$ .

Note that this theorem requires that the supremum be attained at the optimal choice  $\hat{c}$ , but not elsewhere.

Our second result builds on [Theorem 1](#) to study dynamic programming problems with continuous and discrete choices. In every period, the agent makes a continuous and discrete choice  $(c', d')$  based on the state variable  $(c, d)$  consisting of the previous period's choices. We denote the set of possible states by  $\Omega$ . The agent may only make choices that satisfy the constraint

$$(c, c', d, d') \in \Gamma.$$

It will be convenient to write

$$\begin{aligned}\Gamma(c, \cdot; d, \cdot) &= \{(c', d') : (c, c', d, d') \in \Gamma\} \\ \Gamma(c, \cdot; d, d') &= \{c' : (c, c', d, d') \in \Gamma\} \\ \Gamma(\cdot, c'; d, d') &= \{c : (c, c', d, d') \in \Gamma\}.\end{aligned}$$

Let us assume that the agent has a feasible choice at every state, i.e.  $\Gamma(c, \cdot, d, \cdot) \subseteq \Omega$  is non-empty for all  $(c, d) \in \Omega$ .

**Problem 1.** Consider the following dynamic programming problem:

$$V(c, d) = \sup_{(c', d') \in \Gamma(c, \cdot; d, \cdot)} u(c, c'; d, d') + \beta V(c', d'), \quad (1)$$

where the domain of  $V$  is  $\Omega$ . We assume that  $u(\cdot, c'; d, d')$  and  $u(c, \cdot; d, d')$  are differentiable on  $\text{int}(\Gamma(\cdot, c'; d, d'))$  and  $\text{int}(\Gamma(c, \cdot; d, d'))$ , respectively.<sup>5,6</sup>

There are two sources of non-differentiability in [Problem 1](#). As before, the value function may have downward kinks at states where the agent is indifferent between two discrete choices. In addition, [Problem 1](#) introduces constraints. Hence, the value function may have upward (or downward) kinks at states where the agent makes a boundary choice, but prefers an interior choice at some nearby states. As in [Theorem 1](#), our approach is to focus on differentiability at optimal choices away from boundaries.

We define interior choices as follows. First, the continuous component of the choice  $(c', d')$  must satisfy the standard requirement that it is in the interior of today's feasible set. In addition, a second – more subtle – requirement is necessary. Suppose that the agent plans to choose  $(c'', d'')$  tomorrow after choosing  $(c', d')$  today. If the agent were to change  $c'$  a little bit, then  $(c'', d'')$  might become infeasible. If  $(c'', d'')$  is a particularly good choice for the agent tomorrow, the agent would effectively be constrained to choices today which make  $(c'', d'')$  feasible tomorrow. Therefore, the notion of interior choice must take into account the constraint imposed by the subsequent choice of the agent. We require that  $(c'', d'')$  remains feasible after all sufficiently small changes in  $c'$ .<sup>7</sup> These two considerations lead to the following definition.

**Definition 2.** *The choice  $c'$  is a one-period interior choice with respect to  $(c, c'', d, d', d'')$  if*

- (i)  $c' \in \text{int}(\Gamma(c, \cdot; d, d'))$  and
- (ii)  $c' \in \text{int}(\Gamma(\cdot, c''; d', d''))$ .

We establish that the value function is differentiable at optimal interior choices.

**Theorem 2.** *Suppose  $(\hat{c}', \hat{d}')$  and  $(\hat{c}'', \hat{d}'')$  are optimal choices at states  $(c, d)$  and  $(\hat{c}', \hat{d}')$ , respectively. If  $\hat{c}'$  is a one-period interior choice with respect to  $(c, \hat{c}'', d, \hat{d}', \hat{d}'')$ , then  $V(\cdot, \hat{d}')$  is differentiable at  $\hat{c}'$  and satisfies the first-order condition*

$$-u_{c'}(c, \hat{c}'; d, \hat{d}') = \beta V_c(\hat{c}', \hat{d}') = \beta u_c(\hat{c}', \hat{c}''; \hat{d}', \hat{d}'').$$

---

<sup>5</sup> Since we neither study nor require the existence of optimal policies or value functions, we do not impose conditions such as  $\beta \in (0, 1)$ . In particular, if the value function takes infinite values, then there are no maxima and the conditions for our theorems are violated.

<sup>6</sup> This notation accommodates non-stationary problems. For example, the discrete choice set  $D$  could be constructed as  $D = \cup_{t=0}^{\infty} D_t$ , where each pair of sets  $D_t$  and  $D_{t'}$  is disjoint, and  $\Gamma(c, c'; d, \cdot) \subseteq D_{t+1}$  for all  $d \in D_t$  and all  $c, c' \in C$ .

<sup>7</sup> This second condition is somewhat familiar. [Benveniste and Scheinkman \(1979\)](#) require that  $(c, \hat{c}') \in \text{int}(\Gamma)$  where  $\hat{c}'$  is an optimal choice at state  $c$ . Their condition is used to establish a stronger version of sub-differentiability. However, this part of their proof only requires  $c \in \text{int}(\Gamma(\cdot, \hat{c}'))$  which is similar to our second condition.

The optimal one-period interior choice condition of the theorem is unusual as it requires the existence of an optimal continuous choice  $\hat{c}''$  tomorrow (as well as today). Nevertheless, this condition is quite weak in three regards. First, it is only relevant when the problem has constraints; it is automatically satisfied if all continuous choices are made from open sets. Second, the condition does not require the optimal continuous choice tomorrow  $\hat{c}''$  to be an interior choice. Third, the condition does not require the value function to be differentiable at tomorrow's optimal choice (or anywhere else). Even if the agent chooses a kink point tomorrow, the theorem still applies so long as it is feasible for him to change today's choice  $\hat{c}'$  without changing his other choices. Still, the one-period interior choice condition may fail and we explore how to apply [Theorem 2](#) in such cases in [Section 4](#).

A natural extension of [Problem 1](#) is the following version of a dynamic programming problem that incorporates stochastic shocks.

**Problem 2.** *Consider the following stochastic dynamic programming problem:*

$$V(c, d, \theta) = \sup_{(c', d') \in \Gamma(c, \cdot; d, \cdot; \theta)} u(c, c'; d, d'; \theta) + \beta \sum_{\theta' \in \Theta} \pi(\theta' | \theta) V(c', d', \theta'),$$

where the domain of  $V$  is  $\Omega \times \Theta$ . We assume that  $u(\cdot, c'; d, d'; \theta)$  and  $u(c, \cdot; d, d'; \theta)$  are differentiable on  $\text{int}(\Gamma(\cdot, c'; d, d'; \theta))$  and  $\text{int}(\Gamma(c, \cdot; d, d'; \theta))$ , respectively.

The following theorem establishes that the value function is differentiable at optimal choices. It is a stochastic version of [Theorem 2](#).

**Definition 3.** *The choice  $c'$  is a stochastic one-period interior choice with respect to  $(c, c''(\cdot), d, d', d''(\cdot))$  at  $\theta$  if*

- (i)  $c' \in \text{int}(\Gamma(c, \cdot; d, d'; \theta))$  and
- (ii)  $c' \in \text{int}(\Gamma(\cdot, c''(\theta'); d', d''(\theta'); \theta'))$  for all  $\theta'$ .

**Theorem 3.** *Suppose  $(\hat{c}', \hat{d}')$  are optimal choices following  $(c, d, \theta)$  in [Problem 2](#), and  $(\hat{c}''(\cdot), \hat{d}''(\cdot))$  are optimal policies for the following period's choices as a function of  $\theta''$ . If  $\hat{c}'$  is a one-period interior choice with respect to  $(c, \hat{c}''(\cdot), d, \hat{d}', \hat{d}''(\cdot))$ , then  $V(\cdot, \hat{d}')$  is differentiable at  $\hat{c}'$  and satisfies the first-order condition*

$$-u_{c'}(c, \hat{c}'; d, \hat{d}'; \theta) = \beta \sum_{\theta'} \pi(\theta' | \theta) V_c(\hat{c}', \hat{d}', \theta') = \beta \sum_{\theta'} \pi(\theta' | \theta) u_c(\hat{c}', \hat{c}''(\theta'); \hat{d}', \hat{d}''(\theta'); \theta').$$

We omit the proof of this theorem, as it is a straightforward generalization of [Theorem 2](#). The main difference is that there is a convex combination of value functions in the Bellman equation, rather than one single value function. This requires a simple generalization of the upcoming [Lemma 2](#) part (iii) to finite sums. Generalizing to continuous random variables would require generalizing this lemma to integrals.

### 3 Proofs

#### 3.1 Classification of Non-Differentiable Points

This section develops a classification of non-differentiable points of functions. We define upward and downward kinks in terms of (Dini) sub- and superderivatives. Then, we show that every non-differentiable point is either an upward or a downward kink and provide a lemma on important algebraic operations.

Intuitively, we would like to define an upward kink as a point where the slope approaching from the left is greater than the slope approaching from the right (see Figure 2a). Downward kinks would have the converse property.

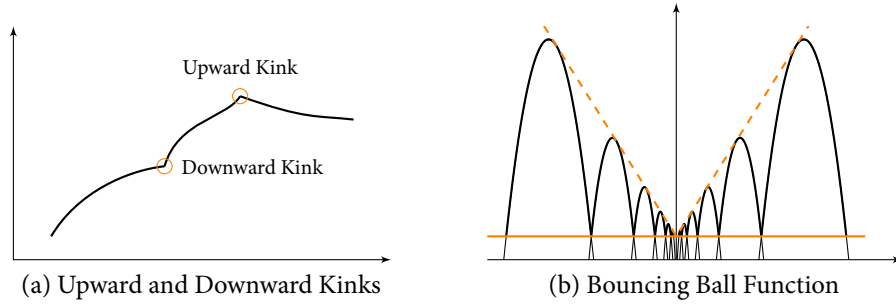


Figure 2: Classifying non-differentiable points

However, we can not use directional derivatives because they may not exist. For instance, consider the bouncing ball function  $F$  depicted in Figure 2b as the upper envelope of a countable set of parabolas  $\{f(\cdot, d)\}_{d \in D}$  where

$$f(c, d) = -\frac{1}{|d|} (c - d) \left( c - \frac{d}{2} \right) \quad \text{and} \quad D = \left\{ \frac{s}{2^n} : s \in \{-1, 1\}, n \in \mathbb{N} \right\}.$$

This function has directional derivatives everywhere except at  $c = 0$ . In particular, the right directional derivative at  $c = 0$ ,

$$\lim_{\Delta c \rightarrow 0^+} \frac{F(\Delta c) - F(0)}{\Delta c}$$

does not exist because the slope oscillates between 0 and  $(\sqrt{2} - 1)^2$ . We resolve this problem by taking limits inferior and superior of the slope, which always exist. According to our

classification,  $c = 0$  is a downward kink but not an upward kink.<sup>8</sup>

**Definition 4.** *The (Dini) sub- and superdifferentials of  $f$  at  $c \in \text{int}(C)$  are*

$$\begin{aligned}\partial_D f(c) &= \left\{ m \in \mathbb{R} : \limsup_{\Delta c \rightarrow 0^-} \frac{f(c + \Delta c) - f(c)}{\Delta c} \leq m \leq \liminf_{\Delta c \rightarrow 0^+} \frac{f(c + \Delta c) - f(c)}{\Delta c} \right\} \\ \partial^D f(c) &= \left\{ m \in \mathbb{R} : \liminf_{\Delta c \rightarrow 0^-} \frac{f(c + \Delta c) - f(c)}{\Delta c} \geq m \geq \limsup_{\Delta c \rightarrow 0^+} \frac{f(c + \Delta c) - f(c)}{\Delta c} \right\}.\end{aligned}$$

If  $\partial_D f(c)$  is non-empty, then we say  $f$  is (Dini) subdifferentiable at  $c$ . Similarly, if  $\partial^D f(c)$  is non-empty, then we say  $f$  is (Dini) superdifferentiable at  $c$ .

**Definition 5.** *If  $f$  is not subdifferentiable at  $c$ , then we say it has an **upward kink** at  $c$ . Similarly, if  $f$  is not superdifferentiable at  $c$ , then we say it has a **downward kink** at  $c$ .*

The following lemma establishes that a non-differentiable point of a function can be classified as either an upward kink or a downward kink.

**Lemma 1** (Differentiability). *A function  $f : \mathbb{R} \rightarrow \mathbb{R}$  is differentiable at  $c$  if and only if  $f$  is both sub- and superdifferentiable at  $c$ . Moreover, if  $f$  is differentiable at  $c$  then  $\{f'(c)\} = \partial_D f(c) = \partial^D f(c)$ .*

*Proof.* The forward direction is straightforward. For the reverse direction, suppose that  $f$  is both sub- and superdifferentiable at  $c$ , so that there exist  $m_* \in \partial_D f(c)$  and  $m^* \in \partial^D f(c)$ . From the definitions,

$$\begin{aligned}\limsup_{\Delta c \rightarrow 0^-} \frac{f(c + \Delta c) - f(c)}{\Delta c} &\leq m_* \leq \liminf_{\Delta c \rightarrow 0^+} \frac{f(c + \Delta c) - f(c)}{\Delta c} \\ \liminf_{\Delta c \rightarrow 0^-} \frac{f(c + \Delta c) - f(c)}{\Delta c} &\geq m^* \geq \limsup_{\Delta c \rightarrow 0^+} \frac{f(c + \Delta c) - f(c)}{\Delta c}.\end{aligned}$$

Since infima are weakly less than suprema, going clockwise, each expression is weakly less than the following one. Therefore, all of the expressions are equal. Thus,  $f$  is differentiable at  $c$  with  $f'(c) = m^* = m_*$ .  $\square$

The following lemma provides some calculus properties of sub- and superdifferentials. Part (iii) provides a sufficient condition for the differentiability of a sum of functions, and plays an important role in the proof of [Theorem 2](#).

---

<sup>8</sup> In similar examples, there are points that are both upward and downward kinks. For example, the function  $f(x) = x \sin \frac{1}{x}$  has an upward and downward kink at  $x = 0$ .



**Lemma 2** (Differential Calculus). *The following statements are true at any  $c$  (along with their superdifferentiable counterparts):*

- (i) *If  $g$  and  $h$  are subdifferentiable, then so is  $g + h$ .*
- (ii)  *$g$  is subdifferentiable if and only if  $-g$  is superdifferentiable.*
- (iii) *If  $g$  and  $h$  are subdifferentiable and  $g + h$  is superdifferentiable, then  $g$ ,  $h$ , and  $g + h$  are differentiable.*

*Proof.* (i) This result follows from the subadditivity property of limits superior that allows us to write

$$\begin{aligned} \limsup_{c \rightarrow 0^-} [g(c) + h(c)] &\leq \limsup_{c \rightarrow 0^-} \left[ g(c) + \limsup_{c \rightarrow 0^-} h(c) \right] \\ &= \limsup_{c \rightarrow 0^-} g(c) + \limsup_{c \rightarrow 0^-} h(c), \end{aligned}$$

and the analogous right limit inferior inequality.

- (ii) Trivial.
- (iii) From part (i),  $g + h$  is subdifferentiable, and hence differentiable by [Lemma 1](#). From part (ii),  $-g$  is superdifferentiable, and part (i) implies  $h = (g + h) + (-g)$  is superdifferentiable. Therefore, [Lemma 1](#) implies  $h$  is differentiable. □

**History:** The notions of a Dini sub- and superdifferentials of one-dimensional functions are special cases of Fréchet sub- and superdifferentials of functions on Banach spaces. For simplicity, the body of our paper uses Dini sub- and superdifferentials (generalizations of all results are in [Appendix A](#)). However, we will discuss the history here in terms of the non-smooth analysis literature which focuses on Fréchet sub- and superdifferentials.

The notions of Fréchet sub- and superdifferentials generalize classical notions from convex analysis to non-convex functions. However, according to [Kruger \(2003\)](#), previous work in mathematics has not applied these concepts because of “rather poor calculus” as  $\partial_F(f + g)(x) \neq \partial_F f(x) + \partial_F g(x)$ . Our approach appears to be novel: we simultaneously study sub- and superdifferentiability of functions to establish full differentiability.

The notions of Fréchet sub- and superdifferentials defined in [Appendix A](#) are standard, and appear in [Schiotzek \(2007, Chapter 9\)](#), although Fréchet superdifferentials only appear in a two-page section on Hamilton-Jacobi equations. The special case of Dini sub-

and superdifferentials is non-standard. The pioneering papers that lead to these definitions are [Clarke \(1975, 1976\)](#), [Penot \(1974, 1978\)](#), and [Bazaraa, Goode and Nashed \(1974\)](#). [Lemma A.1](#) – the Banach space version of [Lemma 1](#) – appears without proof as Proposition 1.3 in [Kruger \(2003\)](#), but does not appear in Schirotzek. Parts (i) and (ii) of [Lemma A.2](#) – the Banach space version of [Lemma 2](#) – appear without proof in the discussion on pages 172 and 181 of Schirotzek, respectively. We believe part (iii) is novel.

### 3.2 Proof of Theorem 1

To establish [Theorem 1](#), we prove that

- (i) non-differentiable points are either upward or downward kinks or both ([Lemma 1](#)),
- (ii) optimal choices may not occur at downward kinks ([Lemma 3](#), [Figure 3a](#)), and
- (iii) upper envelopes may not contain upward kinks ([Lemma 4](#), [Figure 3b](#)).

**Lemma 3.** *If  $\hat{c} \in \text{int}(C)$  is a maximum of  $g : \mathbb{R} \rightarrow \mathbb{R}$ , then  $g$  is superdifferentiable at  $\hat{c}$  with  $0 \in \partial^D g(\hat{c})$ .*

*Proof.* Since  $\hat{c}$  is a maximum, the slope on the left is weakly positive, and the slope on the right is weakly negative. In other words, for any  $\Delta c > 0$ ,

$$\frac{g(\hat{c} - \Delta c) - g(c)}{-\Delta c} \geq 0 \geq \frac{g(\hat{c} + \Delta c) - g(c)}{\Delta c}.$$

Taking limits gives

$$\liminf_{\Delta c \rightarrow 0^-} \frac{g(\hat{c} + \Delta c) - g(c)}{\Delta c} \geq 0 \geq \limsup_{\Delta c \rightarrow 0^+} \frac{g(\hat{c} + \Delta c) - g(c)}{\Delta c},$$

which establishes  $0 \in \partial^D g(\hat{c})$ . □

**Lemma 4.** *If  $F$  is the upper envelope of a (possibly infinite) set of differentiable functions  $\{f(\cdot, d)\}$ , and  $c \in \text{int}(C)$ , and  $F(c) = f(c, \hat{d})$ , then  $F$  is subdifferentiable at  $c$  with  $f_c(c, \hat{d}) \in \partial_D F(c)$ .*

*Proof.* Since  $\hat{d}$  is an optimal choice at  $c$  but (perhaps) not at  $c + \Delta c$ ,

$$f(c + \Delta c, \hat{d}) - f(c, \hat{d}) \leq F(c + \Delta c) - F(c).$$

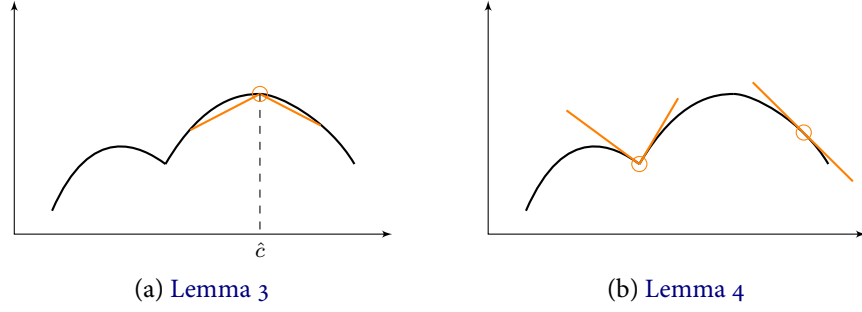


Figure 3: Illustration of Proof of Theorem 1

Dividing by  $\Delta c > 0$ , and taking limits gives

$$f_c(c, \hat{d}) \leq \liminf_{\Delta c \rightarrow 0^+} \frac{F(c + \Delta c) - F(c)}{\Delta c}.$$

Similarly, dividing by  $\Delta c < 0$  and taking limits gives

$$f_c(c, \hat{d}) \geq \limsup_{\Delta c \rightarrow 0^-} \frac{F(c + \Delta c) - F(c)}{\Delta c}.$$

Therefore,  $f_c(c, \hat{d}) \in \partial_D F(c)$ . □

We are ready now to prove Theorem 1 which is restated here.

**Theorem 1.** Suppose  $F$  is the upper envelope of a (possibly infinite) set of differentiable functions  $\{f(\cdot, d)\}$ . If  $(\hat{c}, \hat{d})$  maximizes  $f$ , and  $\hat{c} \in \text{int}(C)$ , then  $F$  is differentiable at  $\hat{c}$  and satisfies the first-order condition  $F'(\hat{c}) = f_c(\hat{c}, \hat{d}) = 0$ .

*Proof.* Lemmata 3 and 4 establish that  $F$  is super- and subdifferentiable at  $\hat{c}$  with  $0 \in \partial^D F(\hat{c})$  and  $f_c(\hat{c}, \hat{d}) \in \partial_D F(\hat{c})$ . Applying Lemma 1, we conclude that  $F$  is differentiable at  $\hat{c}$  with  $F'(\hat{c}) = 0 = f_c(\hat{c}, \hat{d})$ . □

**History:** We sketch the history of the proof steps (i)–(iii). Mirman and Zilcha (1975, Lemma 1) introduced (iii) in the context of a growth model. Instead of using (ii), they ensure that there are no downward kinks by assuming that the objective is jointly concave in all choices. Their proofs are based on directional derivatives, which exist everywhere on concave functions.<sup>9</sup> Benveniste and Scheinkman (1979) generalize their theorem.

<sup>9</sup> Rockafellar (1970, Theorem 23.1) proves the existence of directional derivatives of concave functions, which he traces as far back as Stolz (1893, Satz 10, p. 35) who describes it as a standard result from geometry.

Amir *et al.* (1991, Lemmata 3.3 and 3.4) introduced the proof strategy of (i)–(iii), also in the context of a growth model. To ensure that directional derivatives exist in step (i), they impose a supermodularity assumption on the underlying objective function.

Milgrom and Segal (2002, Corollary 2) were the first to notice that this logic applies without any topological or monotonicity assumptions on the discrete choice set  $D$ . To ensure that directional derivatives exist in step (i), they assumed that  $\{f(\cdot, d)\}_{d \in D}$  is an equidifferentiable set of functions.<sup>10</sup> Their Theorem 3 generalizes Clarke (1975, Theorem 2.1), which in turn generalizes Danskin (1966, Theorem 1).

### 3.3 Proof of Theorem 2

In this section, we prove Theorem 2 which is restated here:

**Theorem 2.** *Suppose  $(\hat{c}', \hat{d}')$  and  $(\hat{c}'', \hat{d}'')$  are optimal choices at states  $(c, d)$  and  $(\hat{c}', \hat{d}')$ , respectively. If  $\hat{c}'$  is a one-period interior choice with respect to  $(c, \hat{c}'', d, \hat{d}', \hat{d}'')$ , then  $V(\cdot, \hat{d}')$  is differentiable at  $\hat{c}'$  and satisfies the first-order condition*

$$-u_{c'}(c, \hat{c}'; d, \hat{d}') = \beta V_c(\hat{c}', \hat{d}') = \beta u_c(\hat{c}', \hat{c}''; \hat{d}', \hat{d}'').$$

The following lemma implies that  $V(\cdot, \hat{d}')$  is subdifferentiable at the optimal choice  $\hat{c}'$  when  $\hat{c}' \in \text{int}(\Gamma(\cdot, \hat{c}''; \hat{d}', \hat{d}''))$ . In other words, upward kinks may only arise when a constraint binds on today's choice; upward kinks in the value function at future dates do not propagate backwards. Note that the lemma is written with different timing, and is applicable in a more general setting than the theorem.

**Lemma 5.** *Suppose  $(\hat{c}', \hat{d}')$  are optimal choices given  $(c, d)$  in Problem 1. If  $c \in \text{int}(\Gamma(\cdot, \hat{c}'; d, \hat{d}'))$ , then the value function  $V(\cdot, d)$  is subdifferentiable at  $c$  with  $u_c(c, \hat{c}'; d, \hat{d}') \in \partial_D V(c, d)$ .*

*Proof.* For all  $c + \Delta c \in \Gamma(\cdot, \hat{c}'; d, \hat{d}')$ , we have the inequality

$$\begin{aligned} & V(c + \Delta c, d) - V(c, d) \\ & \leq \left[ u(c + \Delta c, \hat{c}'; d, \hat{d}') + \beta V(\hat{c}', \hat{d}') \right] - \left[ u(c, \hat{c}'; d, \hat{d}') + \beta V(\hat{c}', \hat{d}') \right] \\ & = u(c + \Delta c, \hat{c}'; d, \hat{d}') - u(c, \hat{c}'; d, \hat{d}'). \end{aligned}$$

Since  $c \in \text{int}(\Gamma(\cdot, \hat{c}'; d, \hat{d}'))$ , this inequality holds for all  $\Delta c$  in an open neighborhood of 0. Dividing both sides by  $\Delta c > 0$  and taking limits gives

$$\liminf_{\Delta c \rightarrow 0^+} \frac{V(c + \Delta c, d) - V(c, d)}{\Delta c} \leq u_c(c, \hat{c}'; d, \hat{d}').$$

---

<sup>10</sup> A set of functions  $\{f(c, \cdot)\}_{d \in D}$  is equidifferentiable at  $c$  if  $[f(c + \Delta c, d) - f(c, d)] / \Delta c$  converges uniformly as  $\Delta c \rightarrow 0$ .

Similarly, dividing both sides by  $\Delta c < 0$  and taking limits gives

$$\limsup_{\Delta c \rightarrow 0^-} \frac{V(c + \Delta c, d) - V(c, d)}{\Delta c} \geq u_c(c, \hat{c}'; d, \hat{d}').$$

Therefore,  $V(\cdot, d)$  is subdifferentiable at  $c$  with  $u_c(c, \hat{c}'; d, \hat{d}') \in \partial_D V(c, d)$ .  $\square$

It remains to show that  $V(\cdot, \hat{d}')$  is superdifferentiable at the optimal choice  $\hat{c}'$ . The Bellman equation in [Problem 1](#) may be decomposed into the recursive equations

$$v(c'; c, d) = \sup_{d' \in \Gamma(c, c'; d, \cdot)} u(c, c'; d, d') + \beta V(c', d') \quad (2a)$$

$$V(c, d) = \sup_{c' \in C} v(c'; c, d) \quad (2b)$$

$$\text{s.t. } c' \in \Gamma(c, \cdot; d, d') \text{ for some } d' \in D.$$

Our approach is to strip away the operations on the right side of (2a) until we arrive at  $V$ , showing that each expression is superdifferentiable at each step. Surprisingly, the subdifferentiability of  $V(\cdot, \hat{d}')$  established above plays a key role.

Since [Theorem 2](#) requires that the optimal choice  $\hat{c}'$  lies in  $\text{int}(\Gamma(c, \cdot; d, \hat{d}'))$ , [Lemma 3](#) implies that  $v(\cdot; c, d)$  is superdifferentiable at  $\hat{c}'$ , and 0 is a superderivative.

The right side of (2a) can be written as

$$G(c') = \sup_{d' \in \Gamma(c, c'; d, \cdot)} g(c', d') \quad (3a)$$

$$g(c', d') = u(c, c'; d, d') + \beta V(c', d'). \quad (3b)$$

Part (i) of the following lemma establishes that  $g(\cdot, \hat{d}')$  is also superdifferentiable at  $\hat{c}'$  with a superderivative of 0. The lemma applies to general static optimization problems and may be applied by setting  $f = g$  and  $F = G$ .

**Lemma 6.** *Suppose  $F$  is the upper envelope of a set of functions  $\{f(\cdot, d)\}$ , and that  $F(c^*) = f(c^*, d^*)$ .*

- (i) *If  $F$  is superdifferentiable at  $c^*$ , then  $f(\cdot, d^*)$  is also superdifferentiable at  $c^*$  with  $\partial^D f(c^*, d^*) \supseteq \partial^D F(c^*)$ .*
- (ii) *If  $f(\cdot, d^*)$  is subdifferentiable at  $c^*$ , then  $F$  is also subdifferentiable at  $c^*$  with  $\partial_D f(c^*, d^*) \subseteq \partial_D F(c^*)$ .*

*Proof.* We only present the proof of part (i), as the proof for part (ii) is analogous. Since  $F$  is superdifferentiable at  $c^*$ , there is some slope  $m^* \in \partial^D F(c^*)$  with

$$\liminf_{\Delta c \rightarrow 0^-} \frac{F(c^* + \Delta c) - F(c^*)}{\Delta c} \geq m^* \geq \limsup_{\Delta c \rightarrow 0^+} \frac{F(c^* + \Delta c) - F(c^*)}{\Delta c}.$$

Since  $F(c) \geq f(c, d^*)$ , we know that

$$F(c^* + \Delta c) - F(c^*) \geq f(c^* + \Delta c, d^*) - f(c^*, d^*).$$

Dividing by  $\Delta c > 0$  and taking limits, we find that

$$m^* \geq \limsup_{\Delta c \rightarrow 0^+} \frac{F(c^* + \Delta c) - F(c^*)}{\Delta c} \geq \limsup_{\Delta c \rightarrow 0^+} \frac{f(c^* + \Delta c, d^*) - f(c^*, d^*)}{\Delta c}.$$

Along with the analogous inequality on the left, this establishes  $m^* \in \partial^D f(c^*, d^*)$ .  $\square$

So far, we have established that  $g(\cdot, \hat{d}')$  is superdifferentiable at  $\hat{c}'$  and that each term in its sum is subdifferentiable. Therefore, [Lemma 2](#) part (iii) implies  $g(\cdot, \hat{d}')$  and  $V(\cdot, \hat{d}')$  are differentiable at  $\hat{c}'$ . We also established that 0 is a superderivative of  $g(\cdot, \hat{d}')$  and  $u_c(\hat{c}', \hat{c}''; \hat{d}', \hat{d}'')$  is a subderivative of  $V(\cdot, \hat{d}')$ , so these are in fact the derivatives. The equality of [Theorem 2](#) follows, and this completes the proof.

**History:** [Lemma 5](#) is a straightforward generalization of [Lemma 4](#), whose history is discussed above. The early envelope theorems for dynamic programming problems ([Mirman and Zilcha \(1975, Lemma 1\)](#) and [Benveniste and Scheinkman \(1979\)](#)) imposed concavity assumptions to establish a form of superdifferentiability to complete the proof. In particular, part (ii) of [Lemma 6](#) – which we did not use to prove the theorem – is reminiscent of [Benveniste and Scheinkman \(1979, Lemma 1\)](#). The proof of [Amir et al. \(1991, Lemma 3.4\)](#) has a similar structure to our [Theorem 1](#). However, in their setting the flow value is differentiable, so their proof is simpler.

## 4 Applications

To illustrate how broadly our theorems may be applied, we present two examples. The first is a classical dynamic programming problem with binary labor choice. This is a straightforward application of [Theorem 2](#). The second is a capital adjustment problem with fixed costs. In this application, the optimal one-period interior choice condition fails, but nevertheless [Theorem 2](#) may be applied.

**Binary Labor Choice:** Consider the following dynamic programming problem with consumption, savings, and a discrete labor choice:

$$\begin{aligned} W(a) &= \max_{(c, a', \ell) \in \mathbb{R}^3} u(c, \ell) + \beta W(a') \\ \text{s.t. } & c \geq 0, a' \geq 0, \ell \in \{0, 1\}, \\ & c + a' = Ra + w\ell, \end{aligned} \tag{4}$$

where  $u(\cdot, \ell)$  is differentiable and  $\beta \in (0, 1)$ . To apply [Theorem 2](#), we reformulate the problem as follows:

$$\begin{aligned} \tilde{W}(a, L) &= \max_{(a', L') \in \Gamma(a, L, \cdot)} u(Ra + wL' - a', L') + \beta \tilde{W}(a', L'), \\ \text{where } \Gamma &= \{(a, a'; L, L') : (a, L) \in \Omega, (a', L') \in \Omega, Ra + wL' - a' \geq 0\} \\ \text{and } \Omega &= [0, \infty) \times \{0, 1\}. \end{aligned}$$

Note the abuse of notation:  $L' = \ell$  is the labor supplied today. This means that  $\tilde{W}$  does not depend on its second argument  $\ell$ , (i.e.  $\tilde{W}(\cdot, L) = W$  for all  $L$ ) as yesterday's labor choice is not pay-off relevant today.

Suppose  $\hat{L}$ ,  $\hat{a}'$ , and  $\hat{c}$  are optimal choices given  $a$ , and that  $\hat{c}'$  is an optimal choice given  $\hat{a}'$ .<sup>11</sup> The optimal one-period interior choice condition of [Theorem 2](#) is that (i)  $\hat{c} > 0$  and  $\hat{a}' > 0$ , and (ii)  $\hat{c}' > 0$ . This condition is very weak in the context of this problem. If the agent's preferences satisfy the Inada condition that the marginal utility of consumption at  $c = 0$  is infinite, then  $\hat{c} > 0$  and  $\hat{c}' > 0$  are satisfied, leaving only the standard requirement that the savings choice  $\hat{a}'$  must be interior. If this condition is met, then the theorem establishes that  $\tilde{W}(\cdot, \hat{L}')$  is differentiable at  $\hat{a}'$ , and the first-order condition

$$u_c(\hat{c}, \hat{L}') = \frac{\beta}{R} \tilde{W}_{a'}(\hat{a}', \hat{L}') = \frac{\beta}{R} W'(\hat{a}')$$

is satisfied.

To summarize,  $W$  exhibits two types of kinks: those arising at savings choices where the agent is indifferent between working or not and those arising where constraints change from non-binding to binding. Kinks at indifference points can not be optimal choices and are therefore irrelevant. Kinks at constraint boundaries are only relevant when tomorrow's choice may become infeasible after a small change in today's choice. The lower bound of saving nothing tomorrow is irrelevant, because saving nothing is feasible regardless of today's choices. The upper bound of saving everything is also irrelevant when the agent's preferences satisfy the Inada condition, because consuming nothing is suboptimal. Therefore, neither type of kink interferes with the application of first-order conditions.

---

<sup>11</sup>  $\hat{c}$  and  $\hat{c}'$  are short-hand for  $\hat{c} = R\hat{a} + w\hat{L}' - \hat{a}'$  and  $\hat{c}' = R\hat{a}' + w\hat{L}'' - \hat{a}''$ .

**Fixed Costs of Capital Adjustment:** In many markets, there are fixed costs associated with adjusting capital stocks. For example, expanding office space involves searching for a new building, transporting furniture, and so on. [Khan and Thomas \(2008\)](#) and [Bachmann, Caballero and Engel \(2006\)](#) study the impact of fixed costs on investment over the business cycle. They apply the envelope theorem of [Benveniste and Scheinkman \(1979\)](#) to the benchmark model of no fixed costs but fall back on numerical methods in the general case. We show that the value function is differentiable at all optimal capital adjustment levels. More generally, this application illustrates how to analyze problems in which the optimal one-period interior choice condition of [Theorem 2](#) fails. The techniques explored here are also applicable to more general adjustment costs, as well as irreversible investment, and problems with bid-ask spreads.

A firm has access to a production technology that allows it to use  $k$  units of capital to produce  $f(k)$  units of output, where  $f$  is differentiable. The market price of the output is normalized to 1. The capital stock  $k$  depreciates at rate  $\delta$ . The firm may adjust the capital stock at any time, but this requires a fixed cost of  $c$  units of output. If it decides to pay this fixed cost, then it may buy or sell units of capital at a price of  $p_k$ . The firm discounts future profits at rate  $\beta$ , and has the following dynamic programming problem:

$$W(k) = \max \begin{cases} f(k) + \beta W((1 - \delta)k), \\ \max_{k' \geq 0} f(k) - c - p_k \left( \frac{k'}{1 - \delta} - k \right) + \beta W(k'). \end{cases}$$

We assume that there is some return to investment and that the firm prefers not to allow the capital stock to depreciate to nothing, i.e.

$$-c + \max_k \sum_{t=0}^{\infty} \beta^t f((1 - \delta)^t k) > \sum_{t=0}^{\infty} \beta^t f(0).$$

The value function  $W$  has downward kinks at capital levels  $k$  where the firm is indifferent between making an adjustment or not. Moreover, when the firm changes from non-adjustment to adjustment (in either direction), it pays a fixed cost which might cause a downward jump in its profits. This could potentially lead to upward kinks in the value function. Below, we apply [Theorem 2](#) to establish: if  $\hat{k}' > 0$  is an optimal choice given  $k$  that involves an adjustment (i.e.  $\hat{k}' \neq (1 - \delta)k$ ), then the value function  $W$  is differentiable at  $\hat{k}'$  and satisfies the first-order condition

$$\frac{p_k}{1 - \delta} = \beta W'(\hat{k}').$$

The firm sets the marginal cost of adjustment equal to the marginal future benefit. Since neither depend on the prior capital stock, there is no history dependence in the capital stock



choice once an adjustment decision has been made. Therefore, the firm repeats through finite cycles in which the capital depreciates and is replenished periodically according to the Euler equation.

We reformulate the problem into the notation of [Problem 1](#):

$$\tilde{W}(k, a) = \begin{cases} \max_{a' \in \{0,1\}} f(k) + \beta \tilde{W}[(1 - \delta)k, a'] & \text{if } a = 0, \\ \max_{k' \geq 0, a' \in \{0,1\}} f(k) - c - p_k \left( \frac{k'}{1 - \delta} - k \right) + \beta \tilde{W}(k', a') & \text{if } a = 1. \end{cases}$$

The optimal one-period interior choice condition of [Theorem 2](#) is satisfied if (i)  $a = 1$  and  $\hat{k}' > 0$ , and (ii)  $\hat{a}' = 1$ . In other words, [Theorem 2](#) applies if the agent makes two adjustments in a row. However, the condition is violated if there is no adjustment in the following period, because a small change in the capital level  $\hat{k}'$  chosen today would imply a small change in the capital level  $(1 - \delta)\hat{k}'$  “chosen” tomorrow. Nevertheless, we establish that  $W$  is differentiable at any optimal capital choice  $\hat{k}' > 0$ .

In the case that the firm waits before readjusting, we may still apply [Theorem 2](#) by bundling the waiting periods together with the first adjustment period into one single period. If the firm waits  $n$  periods before readjusting, the optimal capital level  $\hat{k}'$  maximizes<sup>12</sup>

$$f(k) - c - p_k \left( \frac{k'}{1 - \delta} - k \right) + \beta f(k') + \dots + \beta^n f((1 - \delta)^{n-1} k') + \beta^{n+1} \tilde{W}((1 - \delta)^n k', 1).$$

Equivalently, the optimal capital level  $n$  periods into the future  $\hat{K}'$  maximizes

$$f(k) - c - p_k \left[ \frac{K'}{(1 - \delta)^{n+1}} - k \right] + \beta f\left(\frac{K'}{(1 - \delta)^n}\right) + \dots + \beta^n f\left(\frac{K'}{(1 - \delta)^1}\right) + \beta^{n+1} \tilde{W}(K', 1).$$

We may reinterpret [Theorem 2](#) by treating the flow utility functions as all of the terms before the continuation value.<sup>13</sup> Now, the optimal one-period interior choice condition is that (i)  $a = 1$  and  $\hat{K}' > 0$ , and (ii) an adjustment will be made in the reformulated “tomorrow.” By construction, adjustments are made in both periods, so this condition is met. Therefore, [Theorem 2](#) implies that  $\tilde{W}(\cdot, 1)$  is differentiable at  $\hat{K}' = (1 - \delta)^{n-1} \hat{k}'$ .

<sup>12</sup> We assumed earlier that the firm does not find it profitable to allow the capital stock to depreciate to nothing.

<sup>13</sup> This interpretation of the problem is non-stationary, in that the first period has a different flow value function from the subsequent periods. As discussed earlier, [Theorem 2](#) generalizes easily to non-stationary problems, because the discrete choice can include a time index.

Next, we establish that  $W$  is subdifferentiable at  $\hat{k}'$ . The continuation value  $W$  is bounded below by the value function from choosing non-adjustment for  $n$  periods,

$$H(k') = f(k') + \cdots + \beta^n f((1 - \delta)^{n-1} k') + \beta^{n+1} \tilde{W}((1 - \delta)^n k', 1).$$

By construction,  $H(\hat{k}') = W(\hat{k}')$ , and  $H$  inherits differentiability at  $\hat{k}'$  from  $\tilde{W}(\cdot, 1)$  (as established above). Therefore, part (ii) of [Lemma 6](#) implies that  $W$  is subdifferentiable at  $\hat{k}'$ .

Finally, since  $\hat{k}'$  maximizes

$$-p_k \frac{k'}{1 - \delta} + \beta W(k'),$$

this objective is superdifferentiable at  $\hat{k}'$  by [Lemma 3](#). Moreover, each term is subdifferentiable, so the objective is differentiable by [Lemma 2](#). Thus,  $W$  may be expressed as the difference of two functions that are differentiable at  $\hat{k}'$ . This completes the proof that at any optimal choice  $\hat{k}' > 0$ , the first-order condition

$$\frac{p_k}{1 - \delta} = \beta W'(\hat{k}')$$

is satisfied.

**Numerical Analysis:** Our results may be useful for numerical analysis. [Fella \(2011\)](#) applies our theorems in his generalization of the endogenous grid method of [Carroll \(2006\)](#). He finds his method is substantially faster and more accurate than discretization methods.

## A Banach Space Version

For many dynamic programming problems, there are several continuous choices (we already accommodated arbitrary “discrete” choice spaces above). We generalize our concepts and results to multidimensional spaces, which we number in the same way as in the main text for ease of reference.

Let  $(X, \|\cdot\|)$  be a Banach space (for example,  $X$  could be  $\mathbb{R}^n$ ). We denote

$$X^* = \{\phi : X \rightarrow \mathbb{R} \text{ such that } \phi \text{ is linear and continuous}\}$$

as its topological dual space. The standard notion of differentiability in Banach spaces is due to Fréchet.

**Definition A.2.** A function  $f : X \rightarrow \mathbb{R}$  is **Fréchet differentiable** at  $x$  if there is some  $\phi^* \in X^*$  such that

$$\lim_{\Delta x \rightarrow 0} \frac{f(x + \Delta x) - f(x) - \phi^* \Delta x}{\|\Delta x\|} = 0.$$

$\phi^*$  is called the **Fréchet derivative** of  $f$  at  $x$ , and may be written as  $f'(x)$  or  $f_x(x)$ .

**Theorem A.1.** Suppose  $F$  is the upper envelope of a (possibly infinite) set of Fréchet differentiable functions  $\{f(\cdot, d)\}$ . If  $(\hat{c}, \hat{d})$  maximizes  $f$ , then  $F$  is Fréchet differentiable at  $\hat{c}$  with  $F'(\hat{c}) = f_c(\hat{c}, \hat{d}) = 0$ .

The statement of the generalization of [Theorem 2](#) is identical to the original, apart from the use of Fréchet derivatives.

**Theorem A.2.** Suppose  $(\hat{c}', \hat{c}'', \hat{d}', \hat{d}'')$  are optimal choices following  $(c, d)$  in [Problem 1](#) (in which the utility functions are Fréchet differentiable in the analogous way). If  $\hat{c}'$  is a one-period interior choice with respect to  $(c, \hat{c}'', d, \hat{d}', \hat{d}'')$ , then  $V(\cdot, \hat{d}')$  is Fréchet differentiable at  $\hat{c}'$  and satisfies the first-order condition

$$-u_{c'}(c, \hat{c}'; d, \hat{d}') = \beta V_c(\hat{c}', \hat{d}') = \beta u_c(\hat{c}', \hat{c}''; \hat{d}', \hat{d}''). \quad (5)$$

Notice that the following proofs are shorter than the standard proofs (because we do not have to deal with left and right limits), however this comes at the cost of a loss in economic intuition. We keep the order and numbering similar to the proofs in [Section 3](#) but we omit all the surrounding text, discussing these results.

**Definition A.3.** The **Fréchet subdifferential** of  $f : X \rightarrow \mathbb{R}$  is

$$\partial_F f(x) = \left\{ \phi^* \in X^* : \liminf_{\Delta x \rightarrow 0} \frac{f(x + \Delta x) - f(x) - \phi^* \Delta x}{\|\Delta x\|} \geq 0 \right\},$$

and  $f$  is **Fréchet subdifferentiable** if  $\partial_F f(x)$  is non-empty. Similarly, the **Fréchet super-differential** of  $f$  is

$$\partial^F f(x) = \left\{ \phi^* \in X^* : \limsup_{\Delta x \rightarrow 0} \frac{f(x + \Delta x) - f(x) - \phi^* \Delta x}{\|\Delta x\|} \leq 0 \right\},$$

and  $f$  is **Fréchet superdifferentiable** if  $\partial^F f(x)$  is non-empty.

For completeness, we prove the following standard result which generalizes [Lemma 1](#).

**Lemma A.1.** *A function  $f : X \rightarrow \mathbb{R}$  is Fréchet differentiable if and only if it is both Fréchet sub- and superdifferentiable.*

*Proof.* It is straightforward to show that differentiable functions are sub- and superdifferentiable. Conversely, suppose  $f$  is both Fréchet sub- and superdifferentiable, so that  $\phi_* \in \partial_F f(x)$  and  $\phi^* \in \partial^F f(x)$ . Then from the definitions,

$$\begin{aligned} \liminf_{\Delta x \rightarrow 0} \frac{f(x + \Delta x) - f(x) - \phi_* \Delta x}{\|\Delta x\|} &\geq 0, \\ \limsup_{\Delta x \rightarrow 0} \frac{f(x + \Delta x) - f(x) - \phi^* \Delta x}{\|\Delta x\|} &\leq 0. \end{aligned}$$

The second inequality may be rewritten as

$$\liminf_{\Delta x \rightarrow 0} -\frac{f(x + \Delta x) - f(x) - \phi^* \Delta x}{\|\Delta x\|} \geq 0.$$

From the superadditivity of limits inferior, we deduce

$$\begin{aligned} \liminf_{\Delta x \rightarrow 0} \left[ \frac{f(x + \Delta x) - f(x) - \phi_* \Delta x}{\|\Delta x\|} - \frac{f(x + \Delta x) - f(x) - \phi^* \Delta x}{\|\Delta x\|} \right] &\geq 0 \\ \liminf_{\Delta x \rightarrow 0} [\phi_* - \phi^*] \frac{\Delta x}{\|\Delta x\|} &\geq 0. \end{aligned}$$

But this final equality is only satisfied when  $\phi_* = \phi^*$ . Therefore, the Fréchet sub- and superdifferentials coincide on a singleton, which must be the Fréchet derivative.  $\square$

**Lemma A.2.** *The following statements are true at any  $c$ :*

- (i) *If  $g$  and  $h$  are Fréchet subdifferentiable, then so is  $g + h$ .*
- (ii)  *$g$  is Fréchet subdifferentiable if and only if  $-g$  is Fréchet superdifferentiable.*
- (iii) *If  $g$  and  $h$  are Fréchet subdifferentiable and  $g + h$  is Fréchet superdifferentiable, then  $g$ ,  $h$ , and  $g + h$  are Fréchet differentiable.*

**Lemma A.3.** *If  $\hat{c} \in \text{int}(C)$  is a maximum of  $f : C \rightarrow \mathbb{R}$ , then  $f$  is superdifferentiable at  $\hat{c}$  with  $0 \in \partial^F f(\hat{c})$ .*

*Proof.* Since  $\hat{c}$  is a maximum,  $f(\hat{c} + \Delta c) - f(\hat{c}) \leq 0$  for sufficiently small  $\Delta c \in X$ . Dividing by  $\|\Delta c\|$  and taking limits gives

$$\limsup_{\Delta c \rightarrow 0} \frac{f(\hat{c} + \Delta c) - f(\hat{c})}{\|\Delta c\|} \leq 0.$$

Therefore  $0 \in \partial^F f(\hat{c})$ .  $\square$

**Lemma A.4.** *If  $F$  is the upper envelope of a set of Fréchet differentiable functions  $\{f(\cdot, d)\}$ , and  $c^* \in \text{int}(C)$ , and  $F(c^*) = f(c^*, d^*)$ , then  $F$  is subdifferentiable with  $f_c(c^*, d^*) \in \partial_F F(c^*)$ .*

*Proof.* If  $F(c^*) = f(c^*, d^*)$ , then we have

$$f(c^* + \Delta c, d^*) - f(c^*, d^*) \leq F(c^* + \Delta c) - F(c^*).$$

Subtracting  $\phi \Delta c$ , dividing by  $\|\Delta c\|$ , and taking limits on both sides gives

$$\liminf_{\Delta c \rightarrow 0} \frac{f(c^* + \Delta c, d^*) - f(c^*) - \phi \Delta c}{\|\Delta c\|} \leq \liminf_{\Delta c \rightarrow 0} \frac{F(c^* + \Delta c) - F(c^*) - \phi \Delta c}{\|\Delta c\|}.$$

After setting  $\phi = f_c(c^*, d^*)$ , the left side is zero. Therefore, the right side is non-negative, so  $f_c(c^*, d^*) \in \partial_F F(c^*)$ .  $\square$

**Lemma A.5.** *Suppose  $(\hat{c}', \hat{d}')$  are optimal choices given  $(c, d)$  in [Problem 1](#). If  $c \in \text{int}(\Gamma(\cdot, \hat{c}'; d, \hat{d}'))$ , then the value function  $V(\cdot, d)$  is subdifferentiable at  $c$  with  $u_c(c, \hat{c}'; d, \hat{d}') \in \partial_D V(c, d)$ .*

*Proof.* This proof is omitted; it is straightforward to adapt the proof of [Lemma 5](#) using the technique in the proof of [Lemma 4](#).  $\square$

**Lemma A.6.** *Suppose  $F$  is the upper envelope of the set of functions  $\{f(\cdot, d)\}$ , and that  $F(c^*) = f(c^*, d^*)$ .*

- (i) *If  $F$  is Fréchet superdifferentiable at  $c^*$ , then  $f(\cdot, d^*)$  is also Fréchet superdifferentiable at  $c^*$ .*
- (ii) *If  $f(\cdot, d^*)$  is Fréchet subdifferentiable at  $c^*$ , then  $F$  is also Fréchet subdifferentiable at  $c^*$ .*

*Proof.* We only provide a proof for part (i). Since  $F$  is Fréchet superdifferentiable at  $c^*$ , there is some  $\phi^* \in \partial^F F(c^*)$  with

$$\limsup_{\Delta x \rightarrow 0} \frac{F(x + \Delta x) - F(x) - \phi^* \Delta x}{\|\Delta x\|} \leq 0.$$

Since  $F(c) \geq f(c, d^*)$ , we know that

$$F(c^* + \Delta c) - F(c^*) \geq f(c^* + \Delta c, d^*) - f(c^*, d^*).$$

Subtracting  $\phi^* \Delta c$ , dividing by  $\|\Delta c\|$  and taking limits on both sides yields

$$\limsup_{\Delta x \rightarrow 0} \frac{F(c^* + \Delta c) - F(c^*) - \phi^* \Delta c}{\|\Delta c\|} \geq \limsup_{\Delta x \rightarrow 0} \frac{f(c^* + \Delta c, d^*) - f(c^*, d^*) - \phi^* \Delta c}{\|\Delta c\|}.$$

From the first inequality, the left side is less than 0, which establishes that  $\phi^* \in \partial^F f(c^*, d^*)$ .  $\square$

## References

- AGUIRREGABIRIA, V. and MIRA, P. (2010). Dynamic discrete choice structural models: A survey. *Journal of Econometrics*, **156** (1), 38–67.
- AMIR, R., MIRMAN, L. J. and PERKINS, W. R. (1991). One-sector nonclassical optimal growth: Optimality conditions and comparative dynamics. *International Economic Review*, **32** (3), 625–644.
- BACHMANN, R., CABALLERO, R. J. and ENGEL, E. M. (2006). *Lumpy Investment in Dynamic General Equilibrium*. Working paper, Yale University.
- BAZARAA, M. S., GOODE, J. J. and NASHED, M. Z. (1974). On the cones of tangents with applications to mathematical programming. *Journal of Optimization Theory and Applications*, **13**, 389–426.
- BENVENISTE, L. M. and SCHEINKMAN, J. A. (1979). On the differentiability of the value function in dynamic models of economics. *Econometrica*, **47** (3), 727–732.
- CARROLL, C. D. (2006). The method of endogenous gridpoints for solving dynamic stochastic optimization problem. *Economics Letters*, **91** (3), 312–320.
- CLARKE, F. H. (1975). Generalized gradients and applications. *Transactions of The American Mathematical Society*, **205**, 247–262.
- (1976). A new approach to Lagrange multipliers. *Mathematics of Operations Research*, **1** (2), 165–174.
- DANSKIN, J. M. (1966). The theory of max-min, with applications. *SIAM Journal on Applied Mathematics*, **14** (4), 641–664.
- ECKSTEIN, Z. and WOLPIN, K. I. (1989). The specification and estimation of dynamic stochastic discrete choice models: A survey. *The Journal of Human Resources*, **24** (4), 562–598.
- FELLA, G. (2011). *A Generalized Endogenous Grid Method for Non-Concave Problems*. Working Paper 677, Queen Mary, University of London, School of Economics and Finance.
- KHAN, A. and THOMAS, J. K. (2008). Idiosyncratic shocks and the role of nonconvexities in plant and aggregate investment dynamics. *Econometrica*, **76** (2), 395–436.
- KRUGER, A. Y. (2003). On Fréchet subdifferentials. *Journal of Mathematical Sciences*, **116**, 3325–3358.
- MILGROM, P. and SEGAL, I. (2002). Envelope theorems for arbitrary choice sets. *Econometrica*, **70** (2), 583–601.
- MIRMAN, L. J. and ZILCHA, I. (1975). On optimal growth under uncertainty. *Journal of Economic Theory*, **11** (3), 329–339.
- PENOT, J.-P. (1974). Sous-différentiels de fonctions numériques non convexes. *Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences, Série A*, **278**, 1553–1555.
- (1978). Calcul sous-différentiel et optimisation. *Journal of Functional Analysis*, **27** (2), 248–276.
- RENOU, L. and SCHLAG, K. H. (2009). Ordients: Optimization and comparative statics without utility functions, mimeo.
- RINCÓN-ZAPATERO, J. P. and SANTOS, M. S. (2009). Differentiability of the value function without interiority assumptions. *Journal of Economic Theory*, **144** (5), 1948–1964.
- ROCKAFELLAR, R. T. (1970). *Convex Analysis*. Prince.
- RUST, J. (2008). Dynamic programming. In S. N. Durlauf and L. E. Blume (eds.), *The New Palgrave Dictionary of Economics*, Basingstoke: Palgrave Macmillan.
- SCHIROTZKE, W. (2007). *Nonsmooth Analysis*. Springer.

STOLZ, O. (1893). *Grundzüge der Differential- und Integralrechnung*. Leipzig: B. G. Teubner.